


Cancer detection from stained biopsies using high-speed spectral imaging

EUGENE BROZGOL,^{1,6}  PRAMOD KUMAR,^{1,6} DANIELA NECULA,² IRENA BRONSTEIN-BERGER,¹ MOSHE LINDNER,¹ SHLOMI MEDALION,³ LEE TWITO,¹ YOTAM SHAPIRA,¹ HELENA GONDRA,² IRIS BARSHACK,^{2,4,7} AND YUVAL GARINI^{1,5,7,*}

¹Physics Department and Nanotechnology Institute, Bar Ilan University, Ramat Gan, Israel

²Department of Pathology, Sheba Medical Center, Ramat Gan, Israel

³Gong.io Research, Ramat Gan, Israel

⁴Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

⁵Biomedical Engineering Faculty, Technion – Israel Institute of Technology, Haifa, Israel

⁶Contributed equally

⁷Equal supervision

*yuval.garini@technion.ac.il

Abstract: The escalating demand for diagnosing pathological biopsies requires the procedures to be expedited and automated. The existing imaging systems for measuring biopsies only measure color, and even though a lot of effort is invested in deep learning analysis, there are still serious challenges regarding the performance and validity of the data for the intended medical setting. We developed a system that rapidly acquires spectral images from biopsies, followed by spectral classification algorithms. The spectral information is remarkably more informative than the color information, and leads to very high accuracy in identifying cancer cells, as tested on tens of cancer cases. This was improved even more by using artificial intelligence algorithms that required a rather small training set, indicating the high level of information that exists in the spectral images. The most important spectral differences are observed in the nucleus and they are related to aneuploidy in tumor cells. Rapid spectral imaging measurement therefore can bridge the gap in the machine-aided diagnostics of whole biopsies, thus improving patient care, and expediting the treatment procedure.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Cancer [1] is one of the leading causes of mortality and morbidity; despite modern medical technology, its prevalence is increasing and it affects worldwide health [2]. Cancer diagnostics includes a growing variety of methods ranging from genetic and molecular tests to whole-body imaging. Nevertheless, cancer diagnostics mainly rely on a pathological interpretation of biopsies using traditional glass-slide microscopy of stained tissues and require highly skilled pathologists.

In the last decade, digital pathology [3] (DP) has evolved to meet the growing demand for microscopy-based diagnostics using whole slide imaging (WSI) [4] systems that scan stained slides with high speed and high image quality. It provides pathologists with a screen-based analysis system and other benefits [5]. WSI is based solely on measuring color that provides only three intensities at the red-green-blue (RGB) ranges for each pixel and machine-aided diagnostics of stained biopsies is still under development [4,6,7] without any capability at this time to deal with real medical cases. In current work we show the unprecedented ability of spectral based analysis to achieve cancer diagnostics based on analytical algorithm, or a system based on artificial intelligence (AI) that requires a rather small data base for system training, without demand for large data base for system training. AI is already in use for research pathological

analysis, and the large size of the data required for system training was already highlighted as a severe problem [8].

One of the potential improvements to address this deficiency is to measure spectral images that provides the light spectrum at each pixel of biopsies [9–11]. The spectrum reflects the chemical, biological, and physical state of the substance [6,11,12] and can be used for different biomedical applications [13,14]. Spectral information of cancerous tissues was thoroughly tested before using a large variety of methods [15]. However, there was no consensus on its applicability and most of the work has been devoted to measurements at the tissue level [16] with less attention to the sub-cellular features. A latest work showed the DAPI fluorescence spectral images of colorectal cancer tissue at the sub-cellular level, and excellent classification was achieved [17]. It demonstrates the importance of spectral information of the nucleus for cancer identification. Nevertheless, this method is based on DAPI fluorescence, which takes longer to measure, and it is not the common practice of pathology labs that normally uses haematoxylin and eosin stains for brightfield transmission microscopy. In addition, current spectral imaging (SI) systems have a very long acquisition time for pathological samples [5], which maybe the reason for not using it for cancer diagnostics of biopsies that require the measurement of rather large images.

2. Methods and Materials

Here we present spectral imaging systems for pathological analysis that have very fast whole-slide scanning capability. Except for a scanning microscope stage, the systems contain no moving parts and allow measuring histological slides ‘on the fly’ (Fig. 1). A typical biopsy of 1X1 cm² measured with a 20X magnification results in a spectral image of ~40,000X40,000 pixels with ~40 points in the 400-800 nm spectral range (Fig. 2). Such a measurement takes 5-10 minutes, which can be shortened by using faster cameras.

By measuring tens of cancer cases including breast cancer and colon cancer, we thoroughly studied the spectrum of normal and malignant cells stained with Haematoxylin and Eosin (H&E) and developed algorithms to identify cancer cells; this resulted in diagnostics with very high specificity and sensitivity in all the samples. Moreover, we demonstrated a new spectral imaging modality for digital pathology [18] that provides rapid and accurate diagnostic capabilities. Importantly, this method can be combined with different pathological stains and biomarkers that will provide improved diagnostics and prognostics information.

2.1. Optical architecture of a rapid spectral imaging system

The existing spectral imaging systems [9] are based on one of the following methods:

1. The sample remains still during image acquisition, while some sort of scanning of the wavelength in real or inverse space takes place, and then the sample moves to the next field of view until the whole sample area is measured. These systems normally include filter-based methods that uses either a filter wheel with many narrow single-band filters, a circular or linear variable filter, a liquid crystal tunable filter (LCTF) or acousto optic tunable filter [9].
2. The system measures the spectrum of a single line of the sample (push-broom) and the sample is scanned line-by-line for one stripe of the image, followed by scanning all the stripes to cover the whole sample. The spectral image can also be measured by measuring the spectrum of each pixel, and scanning all the pixels of the image. These methods are often based on a grating, a prism, or a combination of those, including a prism-grating-prism device (PGP).

Both methods suffer from an extremely long measurement time that is not practical for the huge field of view necessary for scanning a pathological sample. In contrast, the newly developed

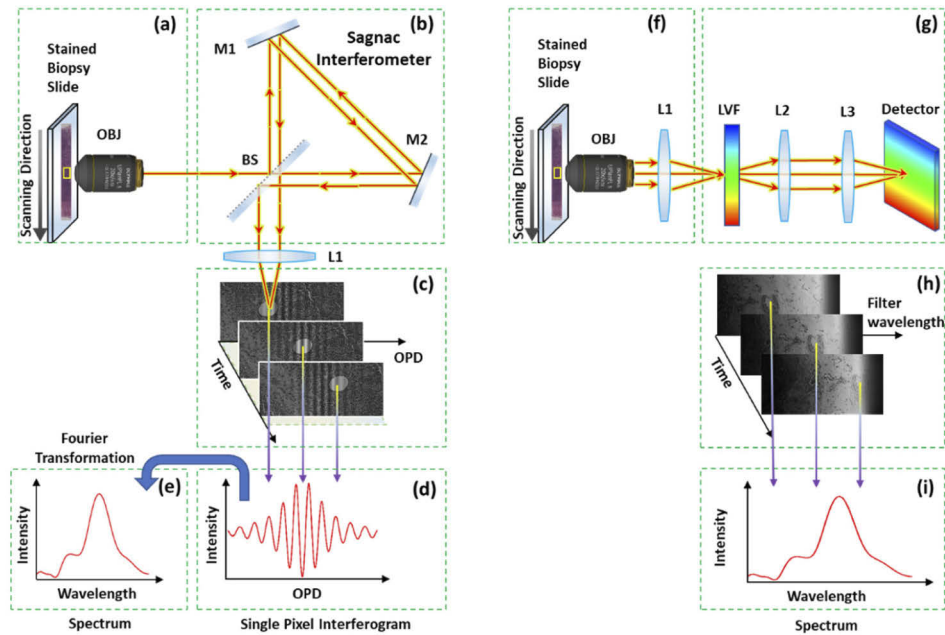


Fig. 1. Schematic diagram of the high-speed spectral imaging system. A Fourier-based system. (a) A stained biopsy slide is scanned at a constant speed ('on the fly') and collimated by an infinity corrected objective lens OBJ. (b) It propagates through a Sagnac common path triangular interferometer that consists of a beamsplitter (BS) and two folding mirrors, M1 and M2. It produces an OPD according to the entrance angle of the beam with respect to the optical axis and focused by lens L1 on the camera. (c) The intensity measured by each pixel along the scanning axis is modified accordingly. Each sample point moves 40-120 pixels, as observed on the camera from one image to another; thus, each image point is measured at 20-50 different OPDs, providing high data acquisition speed. (d) Finally, the interferogram of each pixel is collected from adequate pixels. (e) Each interferogram is Fourier transformed and the whole spectral image is saved. **A LVF-based system.** (f) A stained biopsy slide is scanned at a constant speed ('on the fly'). It is collimated by the objective lens OBJ and focused by the entrance lens L1 on the LVF. (g) The light propagates through the LVF and the telescope (L2, L3) and is measured by the camera. (h). Each entity is measured ~40 times at different positions along the camera. (i) The spectrum of each pixel is collected from adequate pixels from different images along the scan.

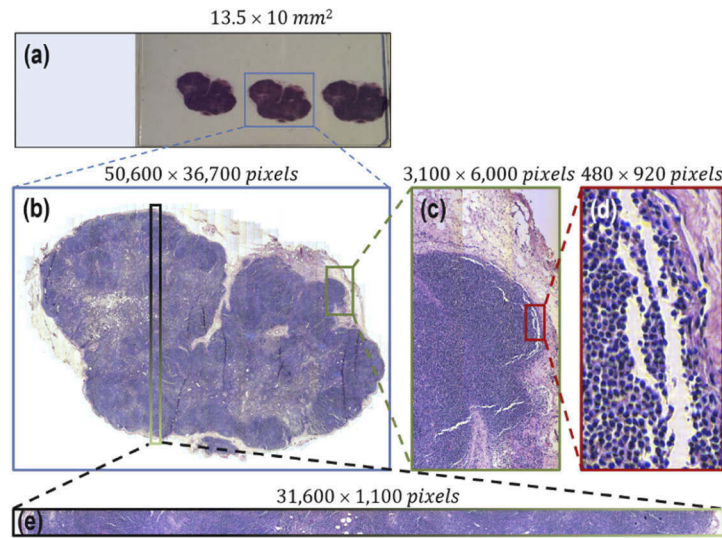


Fig. 2. A rapidly acquired gigapixel spectral image from a histopathological slide. (a) A typical H&E stained biopsy on a microscope slide. The blue box is $13.5 \times 10 \text{ mm}^2$ and the net acquisition took ~ 8 minutes. (b) A white balanced RGB image reconstructed from the full spectral image of $50,600 \times 36,700 \text{ pixels}$ and 40 points in each spectrum in the range of 400–800 nm, which results in a spectral image size of 73 gigapixels. (c) A zoomed section of $3,100 \times 6,000 \text{ pixels}$ (d) Another zoom level image showing $480 \times 920 \text{ pixels}$; it enables one to visualize images at the cellular level. (e) A full stripe from a spectral image that is measured continuously ‘on-the-fly’; it consists of $31,600 \times 1,100 \text{ pixels}$

method [19] provides a very high acquisition speed so that a full biopsy can be measured in a few minutes. The sample is scanned continuously ‘on-the-fly’ (Fig. 1), while images are collected at a high frame rate. In contrast to the push-broom method that measures the sample line by line, here the sample can ‘jump’ 25–120 lines between two consecutive image-captures; thus, it provides high speed and the full spectrum is measured for each sample point. We therefore termed the method a ‘leap-frog’ measurement; the absence of moving optical elements makes the system robust and stable.

For both designs we use a CMOS camera (Lumenera Lt225 NIR, Ottawa, ON, Canada) that has a pixel size of $5.5 \times 5.5 \mu\text{m}^2$ with 2048×1088 pixels. It has 8–12 bits dynamic range, a quantum efficiency of $\sim 55\%$ in the range of 550–600 nm and it can achieve a frame rate of 170 frames per seconds. The systems are connected to an Olympus IX81 microscope side port and we mostly use a 20x objective lens with $\text{NA} = 0.5$ (Olympus UPlanFLN) for sample detection. The combination of the camera pixel size and magnification allows each pixel to view a square of (scan is implemented by using a motorized stage (Prior H117N2IX, Cambridge, UK), which is the only moving part.

The first design is based on Fourier spectroscopy (Fig. 1(a)–(e)). Briefly, a collimated beam produced by the objective lens (Fig. 1(a)) passes a Sagnac interferometer [20–22] (Fig. 1(b)) that splits the light into two different paths; it creates an optical path difference (OPD) between them and merges them again so that they interfere on the detector (Fig. 1(c)). During the scan, the image of each sample point sweeps through different OPDs and its measured intensity changes accordingly (Fig. 1(d)). These intensities are collected for each image point from different pixels measured in different frames and build the interferogram, which is Fourier transformed (FT) to retrieve the spectrum (Fig. 1(e)) [19]. This is a time-consuming operation that somewhat reduces the system’s efficiency.

To overcome this, we developed another leap-frog method achieved with a linear variable filter (LVF) (Fig. 1(f-i)). With respect to the previous method, the unique design provides faster image acquisition (when there is enough light) and immediate post-processing. Here, light from the sample (Fig. 1(f)) is imaged on the LVF and is focused back on the detector (Fig. 1(g)). We use a commercially available LVF (LVVIS, Delta, Hørsholm, Denmark), see [Supplement 1](#) and Fig. 1(s) for system calibration. The system performs well with a spectral accuracy of ~ 1 nm and a spectral resolution of ~ 20 nm.

During the scan, the image of each sample point sweeps through different parts of the LVF and its measured intensity changes accordingly (Fig. 1(h)). These intensities are collected and build the spectrum (Fig. 1(i)).

Spectral imaging systems based on an LVF were already introduced before, but they are mainly based on attaching the LVF directly to the camera array [23,24]. Although such a method has advantages of simplicity and stability, it lacks few important capabilities that we added to our LVF-based system. By adding two cylindrical lenses around the LVF, the light that originates from each pixel can be spread over a line that passes the same wavelength on the LVF, making it immune to elemental defects. In addition, by moving the LVF slightly out of focus, the spectral resolution can be reduced, which will allow a faster acquisition time. Therefore, this setup has unique advantages.

To demonstrate the performance of the systems, especially the required acquisition time and signal to noise, we show a comparison of different spectral imaging methods ([Supplement 1](#) and Table 1s). We analyze the methods according to their fundamental principle and therefore provide the ‘principle-limited’ time for each method. We do not refer to specific commercial systems, as these may not necessarily demonstrate the best possible performance of each method. The comparison demonstrates the high performance of the Fourier-based leap-frog method; this is achieved due to two main reasons: 1. The multiplex advantage of Fourier-based measurements, also known as Fellgett’s advantage, or ‘signal advantage’ [25], and 2. The fact that during the scan, the sample can move tenth of pixels (25-120) in between the measurement of two images. This is in contrast, for example, to the need to scan the sample line-by-line when using a prism or grating.

2.2. Gigapixel spectral image and acquisition speed

The speed of the spectral imaging system in the ‘leap-frog’ method depends on a few parameters [19,21], including the camera frame rate, the exposure time at every frame, and the scan speed. These parameters should be synchronized to achieve an optimal measurement time and should prevent the image from smearing during exposure. We currently measure the biopsies in transmission mode at a frame rate of 150 frames/sec, which can be improved with a higher frame rate camera. We use an exposure time of 10 μ s and a scan speed of 620 μ m/sec; this provides an acquisition time of ~ 5 minutes for a 1X1 cm² biopsy.

A typical spectral image of a biopsy with an area of 13.5X10 mm² measured with a 20X objective lens is shown in Fig. 2. This results in an image size of ~ 50 K X 37K pixels with 40 points in each spectrum and a spectral image size of 73 gigapixels. The sample is measured through 46 stripes (Fig. 2(e)) with a net acquisition time of ~ 8 minutes, an unprecedented speed for such large images. It is somewhat longer than the actual physical limitation due to the limited computer hardware memory in our system.

3. Results

3.1. Spectral properties of normal and cancer nuclei and their origin

To demonstrate the power of the spectral data, we tested the spectra of normal and cancer cells depicted from lymph node biopsies stained with H&E. First, the cells were identified by

a pathologist and marked on an image measured with a standard RGB camera (Fig. 3(a) and Supplement 1). It served as our gold standard for the statistical analysis. The slides were scanned with the spectral imaging system and thousands of nuclei were analyzed by depicting an average spectrum from an area of 9X9 pixels in each nucleus (Fig. 3(b)). Figure 3(c-d) shows the average and standard deviation of spectra from normal (red) and cancer (green) cells as well as a normalized spectrum that emphasizes the spectral-shape differences. Similar differences were found for all eight breast cancer cases (Supplement 1 and Fig. 3s) and another 13 different cases that we studied (Supplement 1, Supplement 1). These differences, both in the intensity and spectral-shape, are instrumental in identifying cancer cells, as shown below.

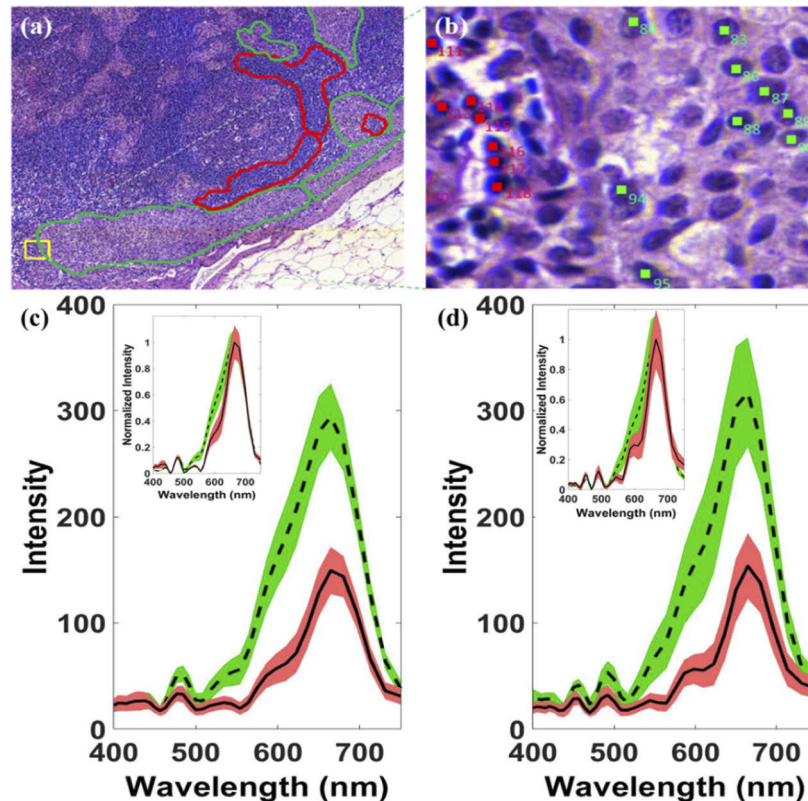


Fig. 3. Evaluating spectral information from H&E stained biopsies. Differentiating cancer from normal cells. (a) An RGB image of $\sim 5\text{ K} \times 6\text{ K}$ pixels, calculated from a spectral image measured from a lymph node biopsy stained with H&E. Some of the cancer and normal areas are denoted (green and red in correspondence). (b) Retrieving normal and cancer spectra from individual nuclei by averaging 9X9 pixels in each nucleus. (c) Spectra peaked from nuclei of normal and breast cancer origin cells from a lymph node biopsy. The mean (black line) and standard deviation (red and green) of the measured spectra from cancer ($N = 855$, dashed line) and normal ($N = 535$, solid line) cells. The insert shows the same spectra normalized to emphasize the difference in the spectral shape. (d) Spectra of intestine origin measured in lymph node biopsies stained with H&E. The mean (black line) and standard deviation (red and green) of the measured spectra from cancer ($N = 580$, dashed line) and normal ($N = 330$, solid line) cells. The insert shows the same spectra normalized to emphasize the difference in the spectral shape.

In order to understand the biochemical origin of the spectral differences in the nucleus of normal and cancer cells, we stained a consecutive set of three tissue sections with Haematoxylin,

Eosin, or both (Supplement 1 and Fig. 19s). The respective absorbance spectra were calculated from the transmission spectra by $A(\lambda) = -\log[I(\lambda)/I_0(\lambda)]$, as shown in Fig. 4(a-c), where $I(\lambda)$ is the measured transmission spectrum and $I_0(\lambda)$ is the background spectrum, measured through an unstained slide. As shown, Eosin absorption is rather weak and similar for both normal and cancer cells (Fig. 4(b)). In contrast, the absorption spectrum of Haematoxylin in normal nuclei is significantly higher, and even more important is the appearance of a well-defined spectral signature at the 540-600 nm wavelengths. This difference in the absorption results in the transmission deep observed in the spectra (Fig. 3(c, d)). Haematoxylin is absorbed by chromatin and its spectra was previously measured [18], but it was not compared in normal and cancerous nuclei. The absorption of haematoxylin (not its spectrum) was also measured and it was found that the dye is not stoichiometric (its intensity is not simply proportional to the DNA content) [26,27]. Our results show that the haematoxylin spectrum provides an excellent signature to distinguish normal cells from cancer ones, as shown below.

Further observation of the spectra shows that the shallow peak in the haematoxylin spectra at the range of 500-550 nm (Fig. 4(a)) also appears in the eosin spectra (Fig. 4(b)). Therefore, in order to obtain an adequate unbiased quantitative analysis of the chromatin content, which is correlated with the cancer grade, it is necessary to eliminate the contribution of the eosin spectrum from the measured spectra. Various methods can be used and we perform this elimination by using a spectral unmixing algorithm, as explained below.

3.2. Extracting DNA content (ploidy) from the spectral data

Breast cancer cells are known to have a larger copy number of chromosomes or elevated chromatin levels (aneuploidy); the correlation between cancer grade and ploidy was shown for different cancer cases [27,28]. Since haematoxylin is absorbed by DNA [26], the absorption of cancer cells is expected to be higher than that of normal cells (Fig. 4(a, c)). Nevertheless, cancer cells seldom have a larger volume, and to estimate the total DNA content in the nucleus, it is necessary to sum up the total absorbance in all the pixels of the nucleus.

In order to obtain a quantitative amount of chromatin, we start by performing a spectral decomposition algorithm that finds the weights of each of the reference spectra $C_i \geq 0$, in our case haematoxylin and eosin [29,30] that minimizes the minimal square error between the measured spectrum $I(\lambda)$ and the reconstructed one, $\arg \min_{C_i \geq 0} \left[I(\lambda) - \sum_i C_i I_i(\lambda) \right]^2$, where $I_i(\lambda)$ are the reference spectra of eosin and haematoxylin [29,30]. It provides the ratio of the absorbance spectrum of both haematoxylin and eosin with respect to the defined spectral references of these two stains in each pixel of the nucleus.

After finding these components, we sum up the haematoxylin absorption values over all the pixels in each nucleus and divide the sum by the total absorption from an average of normal cells:

$$Pl = \frac{\sum_{x,y \in N} C_{x,y}}{\langle \sum_{x,y \in N} C_{x,y} \rangle_{normal}} \quad (.)$$

This value represents the ploidy of the cell. The scatter plot in Fig. 4(d) shows the ploidy distribution of normal and cancer cells for one cancer case. We found an average of $\mu_C = 3.40 \pm 1.0$ for cancer cells ($N = 407$) and $\mu_N = 2.0 \pm 0.6$ ($N = 407$) for normal cells. Similar values were found in all other cases (Supplement 1, Supplement 1). This value indicates the quantitative power of the method and indicates that cancer cells in this lymph node biopsy have a higher ploidy value of about 3.4.

3.3. Spectral cancer detection

To further test the accuracy of cancer cell identification using the spectral information, we tested 35 cases representing eight different origins of cancer, including three of gynecologic origin

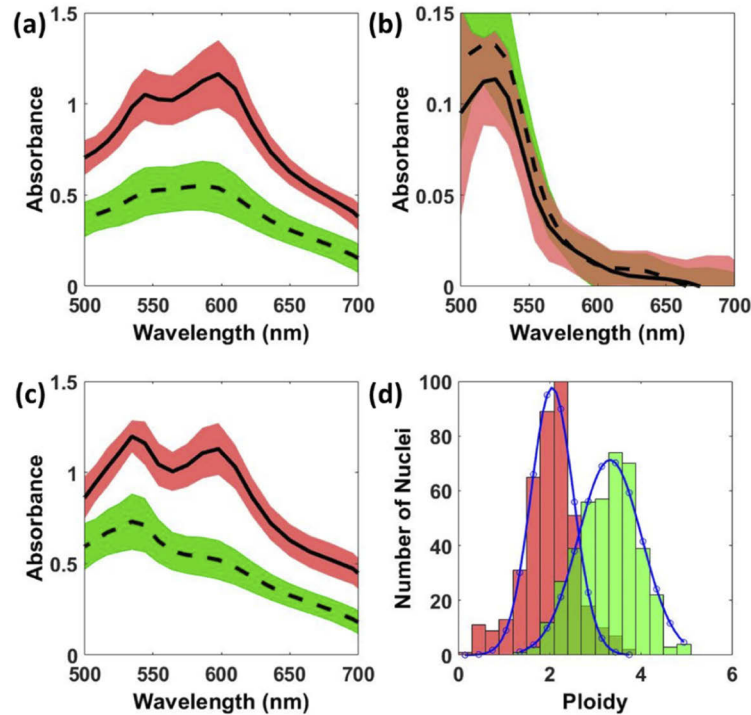


Fig. 4. Absorbance of Haematoxylin and Eosin stains from nuclei of normal and malignant cells. (a) Haematoxylin absorption spectrum of cancer ($N = 50$, dashed line) and normal ($N = 50$, solid line) nuclei. The red and green shades describe the range of the standard deviation of the spectra. (b) Eosin absorption spectra showing relatively low and similar absorption for normal ($N = 50$, solid line) and cancer nuclei ($N = 50$, dash line). (c) Absorption spectra from a biopsy labelled with both H&E. Note the significant difference between cancer ($N = 55$, dashed line) and normal spectra ($N = 55$, solid line). (d) Evaluating the ploidy number for breast cancer. Gaussian fit results in a mean $\mu_N = 2.0 \pm 0.6$ for normal cells ($N = 407$, red) and $\mu_c = 3.40 \pm 1.0$ for cancer cells ($N = 407$, green).

(Endometrium and ovary), two from GI (colon and ampulla), breast, intestinal, and pancreatic cancer (Supplement 1 and Supplement 1).

Biopsies were prepared according to a standard histological procedure for tissue preparation used at the Department of Pathology, Sheba, Tel Hashomer: (a) Formalin fixation; (b) Paraffin-embedding (FFPE); (c) Block sectioning at 4 μm ; (d) Slide staining with H&E in a Leica Autostainer according to the manufacturer's specifications (Leica Biosystems, USA).

3.3.1. Supervised classification scheme for detecting cancer nuclei

We start by demonstrating the accuracy of a classification that considers only the spectral information without taking into account the morphological data. As mentioned before, the 'ground truth' we use is based on the pathologists' diagnosis of cells. We adopted a method based on two parameters, I_1 and I_2 . We first define two reference spectra for normal and cancer cells, $I_{\lambda,RN}$ and $I_{\lambda,RC}$, from a biopsy. Then, we calculate the two parameters: I_1 , the ratio of the mean square error (MSE) of each tested spectrum with the above two spectra, and I_2 , the ratio of

the MSE of each of the normalized spectra with the two normalized reference spectra:

$$I_1 = \frac{\int_{\lambda_1}^{\lambda_2} (I_\lambda - I_{\lambda,RN})^2 d\lambda}{\int_{\lambda_1}^{\lambda_2} (I_\lambda - I_{\lambda,RC})^2 d\lambda} ; I_2 = \frac{\int_{\lambda_1}^{\lambda_2} (\hat{I}_\lambda - \hat{I}_{\lambda,RN})^2 d\lambda}{\int_{\lambda_1}^{\lambda_2} (\hat{I}_\lambda - \hat{I}_{\lambda,RC})^2 d\lambda}$$

where I_λ is the tested pixel spectrum and \hat{I}_λ is a normalized spectrum, $\hat{I}_\lambda = I_\lambda / I_{\lambda_{\max}}$. Figure 5(a-b) shows a scatter plot of the two parameters on a log-log scale. The distinction between normal and cancer cells is excellent and cancer cells can be identified by clustering algorithms [31].

Cells were allocated by a segmentation algorithm and tested according to the pathologist's identification. Use of this classification procedure provides excellent results. For eight breast cancer cases (Table 1) the truth matrix gives 93.7% true positive (TP, N= 370) and 97.4% true negative (TN, N= 290) cells. The results for all cases give 95.4% TP and 98.4% TN, see Supplement 1 for details of more cases.

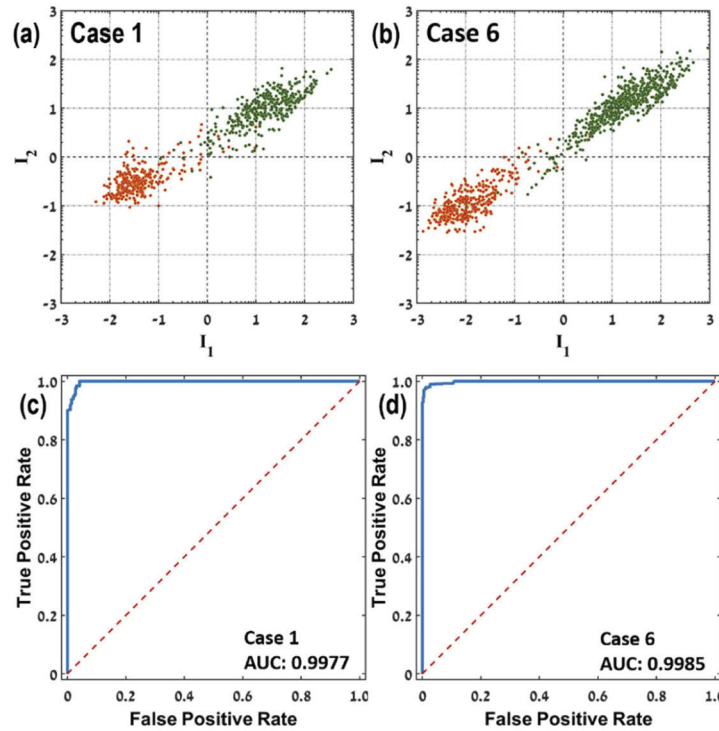


Fig. 5. Spectral-based classification of an H&E stained biopsy. Results for the classification scheme for differentiating cancer from normal cells (see the text). (a) In case 1, 97.8% of the cancer cells were classified as true positive (TP), whereas 96.9% of the normal cells were classified as true negative (TN). (b) For case 6, 94.0% were classified as TP, whereas 100.0% were classified as TN. (c-d) The high Area Under Curve (AUC) as calculated using deep learning algorithm and highlights the pivotal role of spectral information in the classification and the improved classification of cancer cells.

We also tested an iterative scheme that refines the global reference spectra by performing one classification with the global reference spectra, followed by selection of new reference spectra from the biopsy itself. This improves the results by ~3% on average. All cases show high accuracy on a cell-by-cell basis and demonstrate the importance of the spectral information for cancer identification.

Table 1. Confusion matrix showing the percentage of cells for two different cancer cases from breast and gynecological origin as well as the average for all 35 measured cases. For details on number of cells for each case, see Figures S8-S22

Cancer type	True Status	Classification	
		% Normal Cells	% Cancer Cells
Breast	Normal	97.4	2.6
	Cancer	6.3	93.7
Gynecological origin	Normal	99.5	0.5
	Cancer	2.0	98.0
Average: all cases	Normal	98.4	1.6
	Cancer	4.6	95.4

Finally, we tested another improvement analysis to the classification based on the scatter plot for each cancer case by using k-means algorithm, which is also built-in Matlab. Here we find two centroids that belongs to normal and cancer cells that minimize the sum of point-to-centroid distances, summed over the two clusters. K-means can be used for several iterations and we found that it improves the classification by ~4.3%.

3.3.2. Deep learning scheme for detecting cancer nuclei

The above analysis used only the spectral data at each pixel, without taking into account the spatial information at all, which definitely contains important information. In order to incorporate the spatial information into our analysis, we further built an automated algorithm that utilizes the spatial information in addition to the spectral data, using computer vision (CV) approach.

In the last years deep learning algorithms are a main subject for research in the AI field. These systems are widely used for computer vision tasks such as classification, detection and segmentation. Since 2012, Convolutional Neural Networks (CNN) got the best scores in all significant CV competitions with large margin from classical image-processing methods.

The main advantage of deep learning compared to classical image processing and machine learning techniques, is that deep learning algorithms learn complex representations of the data in the latter layers by building up of several simple representations in the first layers of the network. These representations are being learnt by the network itself using an iterative optimization process.

3.4. Data preparation

In order to utilize a deep learning algorithm for enhancing performance of differentiating normal cells from cancer cells, we extracted healthy and cancer regions from the slides, and trained a CNN to classify each image to its correct class.

We split the data into 3 distinct groups: training, validation, and test sets. Each group contained data from distinct slides. Each slide image was divided into 200X200 pixels crops, corresponding to 5-20 nuclei per crop, with a 50% overlap between crops taken from the same slide and each of the crops contained only cancer or normal cells (see [Supplement 1](#)). This way, we ended up with a total of 1442 normal samples and 1991 cancer samples. We normalized each crop by dividing the values in each channel by the maximum value in the same crop's channel.

3.5. System architecture

We used a K-fold cross-validation process in which each fold used separate slides for the training, validation, and test sets; later the model was "re-trained" and each fold was evaluated separately. When splitting the data, we verified that the classes are approximately balanced in each set.

The results were then averaged over all k folds. By using K-fold, we reduced the possibility of randomly picking an “easy” slide for testing our model. For better separation, each of the train/eval/test sets contained crops from different slides, so that the system will not be able to use data extracted from the same slide at prediction time.

We used the Keras [32] framework, an open source code, to build a CNN with a MobileNet architecture [33] (Supplement 1) with randomly initialized weights. We also considered more complex architectures such as ResNet. We focused on the lightweight MobileNet since systems in production are limited in their hardware, and many times does not contain advanced GPUs. For the spectral images, the number of the input channels of the MobileNet was increased to 40, corresponding to the number of channels in the spectral data, so that the input shape was: [Batch Size, Width = 200, Height = 200, Channels = 40]. The output layer dimension was decreased from 1000 to 2, followed by a softmax function, in order to perform a binary classification for the normal/cancer classes and we optimized a categorical cross-entropy loss on these outputs, using an Adam optimizer.

During training, we used a dropout regularization layer right after the fully connected layer. In order to perform a binary classification for the normal/cancer classes, we optimized a binary cross-entropy loss on these outputs, using an Adam optimizer [34].

Feeding our network with the data for several dozen epochs improved the classification performance for the spectral images to an average AUC of 0.999 ± 0.001 , corresponding to 99.1% accuracy over all test crops. Figure 5(c-d) shows an area under the curve that exhibits excellent classification for two breast cancer cases, and for more results see Supplement 1 and Supplement 1.

4. Conclusions and discussion

We described a unique system and method for rapidly measuring and analyzing spectral images of pathological biopsies. The rapid acquisition of a typical $10 \times 10 \text{ mm}^2$ sample takes ~5 minutes and results in an image size of ~50K X 50K pixels with 40 points in the spectral range of 450-800 nm for each pixel with a sampling pixel size of 275 nm. To the best of our knowledge, such a high speed for spectral imaging acquisition was never reported before. It is achieved with a Fourier-based system that measures the sample on-the-fly, and jumps ~25 pixels between the measurement of two consecutive images. Due to the multiplex advantage of the method, it allows to use a very short exposure time of ~20 μs and speedup the measurement.

We tested the system on 35 different neoplastic types and grades of carcinomas and adenocarcinomas that originated from various tissues and identified the spectral signatures of normal and cancer cells according to their nuclear spectra. Spectral signatures of the nuclei enabled us to separate the contribution of haematoxylin and calculate the ploidy number. Further classification algorithms that we developed for identifying cancer cells resulted in unprecedented accuracy at the single cell level.

Our supervised analysis used only the spectra of the normal and cancer cells and yet, the classification already exhibited very high accuracy for cancer detection (>95%). Our study therefore emphasizes the importance of spectral information and its advantage in identifying cancer cells with very high accuracy, even at the single cellular level, and not just at the tissue level, which is easier. By taking into account the morphological features as well, the classification performance can become fully reliable, as demonstrated by using AI, which resulted in an unprecedented accuracy for cancer cell identification (>99%). It therefore has the potential to transform the methodology by which pathological diagnostics are performed.

How does our system perform with respect to other systems? Although it is not the purpose of this manuscript to compare other methods, we note that Ortega et al. have lately performed a comprehensive review of spectral imaging systems for pathological applications [35]. They mainly found pathological studies that uses spectral imaging systems that are based on LCTF,

which is rather slow with respect to our method. Furthermore, they conclude that given the small number of studies and lack of common practices and standards, it is still difficult to assess the capabilities of spectral imaging to pathology. On the other hand, Awan et al. [36] have studied the accuracy of cancer detection when using spectral imaging both in the visible and infrared regimes using two LCTF's. They found excellent classification accuracy for various cancer cases when using both the visible and infrared spectra. This is fully correlated with our results that provides more than 99% accuracy in cancer detection using only the visible spectral range.

In overall, our results demonstrate the efficiency of using spectral imaging for pathological analysis, while we present an affordable system both in terms of its simplicity, measurement time of whole biopsies and usability on normally stained slides.

Future work should focus on analyzing additional parameters, such as the density of cancer cells, their spread, and their proximity to histological features. Finally, we noted that the optical methods of 'leap-frog' and 'wavelength-spread' hold even greater promise for measuring multiple probes of proteins (biomarkers) and genetic features [37], either in bright field or fluorescence; hence, they can become important tools for multiplex labelling of biological samples, which is becoming a crucial tool in research, especially for predicting drug response and precision medicine applications [32–34].

Funding. Bar-Ilan University; Israel Science Foundation (1219/17, 1902/12).

Acknowledgments. We thank Irina Marin from Pathology Department, Sheba Medical Center for her help with biopsies handling.

Disclosures. Bar Ilan University has filed a patent based on this study (inventors: YG, IB, IB). All other authors declare that they have no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer Journal for Clinicians* **61**, 69–90 (2011).
2. S. H. Hassanpour and M. Dehghani, "Review of cancer from perspective of molecular," *Journal of Cancer Research and Practice* **4**(4), 127–129 (2017).
3. R. Bhargava and A. Madabhushi, "Emerging themes in image informatics and molecular analysis for digital pathology," *Annu. Rev. Biomed. Eng.* **18**(1), 387–412 (2016).
4. S. Mukhopadhyay, M. D. Feldman, E. Abels, R. Ashfaq, S. Beltaifa, N. G. Cacciabeve, H. P. Cathro, L. Cheng, K. Cooper, G. E. Dickey, R. M. Gill, R. P. Heaton, R. Kerstens, G. M. Lindberg, R. K. Malhotra, J. W. Mandell, E. D. Manlucu, A. M. Mills, S. E. Mills, C. A. Moskaluk, M. Nelis, D. T. Patil, C. G. Przybycin, J. P. Reynolds, B. P. Rubin, M. H. Saboorian, M. Salicru, M. A. Samols, C. D. Sturgis, K. O. Turner, M. R. Wick, J. Y. Yoon, P. Zhao, and C. R. Taylor, "Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study)," *American Journal of Surgical Pathology* **42**(1), 39–52 (2018).
5. F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: Whole-slide imaging and beyond," *Annu. Rev. Pathol.: Mech. Dis.* **8**(1), 331–359 (2013).
6. W. Liu, L. Wang, J. Liu, J. Yuan, J. Chen, H. Wu, Q. Xiang, G. Yang, and Y. Li, "A comparative performance analysis of multispectral and RGB imaging on HER2 status evaluation for the prediction of breast cancer prognosis," *Translational Oncology* **9**(6), 521–530 (2016).
7. A. B. Farris, C. Cohen, T. E. Rogers, and G. H. Smith, "Whole slide imaging for analytical anatomic pathology and telepathology: Practical applications today, promises, and perils," *Archives of Pathology and Laboratory Medicine* **141**(4), 542–550 (2017).
8. A. Kleppe, O.-J. Skrede, S. de Raedt, K. Liestøl, D. J. Kerr, and H. E. Danielsen, "Designing deep learning studies in cancer diagnostics," *Nat. Rev. Cancer* **21**(3), 199–211 (2021).
9. Y. Garini and E. Tauber, *Spectral Imaging: Methods, Design, and Applications* (Springer, Berlin, Heidelberg, 2013), pp. 111–161.
10. Y. Garini, I. T. Young, and G. McNamara, "Spectral imaging: principles and applications," *Cytometry Part A* **69A**(8), 735–747 (2006).
11. R. M. Levenson and J. R. Mansfield, "Multispectral imaging in biology and medicine: slices of life," *Cytometry Part A* **69A**(8), 748–758 (2006).

12. D. C. Fernandez, R. Bhargava, S. M. Hewitt, and I. W. Levin, "Infrared spectroscopic imaging for histopathologic recognition," *Nat. Biotechnol.* **23**(4), 469–474 (2005).
13. M. J. Baker, H. J. Byrne, J. Chalmers, P. Gardner, R. Goodacre, A. Henderson, S. G. Kazarian, F. L. Martin, J. Moger, N. Stone, and J. Sulé-Suso, "Clinical applications of infrared and Raman spectroscopy: state of play and future challenges," *Analyst* **143**(8), 1735–1757 (2018).
14. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomed. Opt.* **19**(1), 010901 (2014).
15. M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: a review," *IEEE Reviews in Biomedical Engineering* **2**, 147–171 (2009).
16. R. Pourreza-Shahri, F. Saki, N. Kehtarnavaz, P. Leboulluec, and H. Liu, "Classification of ex-vivo breast cancer positive margins measured by hyperspectral imaging," 2013 IEEE International Conference on Image Processing, ICIP 2013 - Proceedings 1408–1412 (2013).
17. K. Liu, S. Lin, S. Zhu, Y. Chen, H. Yin, Z. Li, and Z. Chen, "Hyperspectral microscopy combined with DAPI staining for the identification of hepatic carcinoma cells," *Biomed. Opt. Express* **12**(1), 173 (2021).
18. P. A. Bautista and Y. Yagi, "Digital simulation of staining in histopathology multispectral images: enhancement and linear transformation of spectral transmittance," *J. Biomed. Opt.* **17**(5), 056013 (2012).
19. M. Lindner, Z. Shotan, and Y. Garini, "Rapid microscopy measurement of very large spectral images," *Opt. Express* **24**(9), 9511 (2016).
20. J. Zhao and R. L. McCreery, "Multichannel FT-Raman spectroscopy: noise analysis and performance assessment," *Appl. Spectrosc.* **51**(11), 1687–1697 (1997).
21. A. Barducci, D. Guzzi, C. Lastrì, P. Marcoionni, V. Nardino, and I. Pippi, "Theoretical aspects of Fourier transform spectrometry and common path triangular interferometers," *Opt. Express* **18**(11), 11622 (2010).
22. Y. Garini, M. Yuval, S. Macville, R. A. du Manoir, M. Buckwald, N. Lavi, D. Katzir, I. Wine, E. Bar-Am, D. Schröck, T. Cabib, and Ried, "Spectral karyotyping," *Bioimaging* **4**(2), 65–72 (1996).
23. P. Serruys, A. Sima, S. Livens, B. Delaure, K. Tack, B. Geelen, and A. Lambrechts, "Linear variable filters - A camera system requirement analysis for hyperspectral imaging sensors onboard small Remotely Piloted Aircraft Systems," 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing 2014, 1–4 (2014).
24. I. G. E. Renhorn, D. Bergström, J. Hedborg, D. Letalick, and S. Möller, "High spatial resolution hyperspectral camera based on a linear variable filter," *Optical Engineering* **55**(11), 114105 (2016).
25. R. G. Sellar and G. D. Boreman, "Comparison of relative signal-to-noise ratios of different classes of imaging spectrometer," *Appl. Opt.* **44**(9), 1614–1624 (2005).
26. S. Biesterfeld, S. Beckers, M. del Carmen Villa Cadenas, and M. Schramm, "Feulgen staining remains the gold standard for precise DNA image cytometry," *Anticancer Research* **31**, 53–58 (2011).
27. P. van Loo, S. H. Nordgard, O. C. Lingjaerde, H. G. Russnes, I. H. Rye, W. Sun, V. J. Weigman, P. Marynen, A. Zetterberg, B. Naume, C. M. Perou, A. L. Børresen-Dale, and V. N. Kristensen, "Allele-specific copy number analysis of tumors," *Proc. Natl. Acad. Sci. U. S. A.* **107**(39), 16910–16915 (2010).
28. K. Cyll, E. Ersvaer, L. Vlatkovic, M. Pradhan, W. Kildal, M. Avranden Kjaer, A. Kleppe, T. S. Hveem, B. Carlsen, S. Gill, S. Löffler, E. S. Haug, H. Waehre, P. Sooriakumaran, and H. E. Danielsen, "Tumour heterogeneity poses a significant challenge to cancer biomarker research," *Br. J. Cancer* **117**(3), 367–375 (2017).
29. R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geoscience and Remote Sensing Letters* **12**(1), 38–42 (2015).
30. Y. Garini, N. Katzir, D. Cabib, R. Buckwald, D. Soenksen, and Z. Malik, "Fluorescence imaging spectroscopy and microscopy," *Fluorescence Imaging Spectroscopy and Microscopy* (X. Wang and B. Herman, eds.) (Springer, 1996), 87–124.
31. M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems* **17**(2/3), 107–145 (2001).
32. Team Keras, "Deep Learning for humans," Github (2022), <https://github.com/keras-team/keras>.
33. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861 (2017).
34. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings (2014).
35. S. Ortega, M. Halicek, H. Fabelo, G. M. Callico, and B. Fei, "Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review [Invited]," *Biomed. Opt. Express* **11**(6), 3195–3233 (2020).
36. R. Awan, S. Al-Maadeed, and R. Al-Saady, "Using spectral imaging for the analysis of abnormalities for colorectal cancer: When is it helpful?" *PLOS ONE* **13**(6), e0197431 (2018).
37. S. M. Lewis, M.-L. Asselin-Labat, Q. Nguyen, J. Berthelet, X. Tan, V. C. Wimmer, D. Merino, K. L. Rogers, and S. H. Naik, "Spatial omics and multiplexed imaging to explore cancer biology," *Nat. Methods* **18**(9), 997–1012 (2021).